doi 10.18699/bgrs2024-1.1-19

Systematic analysis of tandem genome amplification events in *Yersinia pestis*

Konanov D.N.[1], Liubimova O.N.[2], Kovrizhnikov A.V.[3], Sonets I.V.[1], Balykova A.N.[3], Lukina-Gronskaya A.V.[1], Speranskaya A.S.[1], Eroshenko G.A.[3], Ilina E.N.[1], Govorun V.M.[1], Kutyrev V.V.[3]

[1] *Research Institute for System Biology and Medicine of Rospotrebnadzor, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Russia*

[3] *Russian Research Anti-Plague Institute "Microbe", Saratov, Russia*

** konanovdmitriy@gmail.com*

Key words: tandem genome amplification; *Yersinia pestis*

Motivation and Aim: Tandem genome amplification is quite common in bacteria growing under stress conditions [1] while spontaneous genome amplification events are rare, not stable and generally poorly described. In most cases, tandem amplification is a result of RecA-dependent recombination which requires short DNA repeats flanking the amplified region. Plague pathogen *Yersinia pestis* is a unique organism that contains an enormous number of short repeat sequences in its genome [2] and as a result is prone to spontaneous genome rearrangements including large tandem duplication events. Now, rearrangement dynamics in *Y. pestis* is considered as one of its main adaptation mechanisms [3], and we suppose that genome amplification events also might play a role in it by affecting gene expression. In this study, we aimed to systematically consider the frequency of spontaneous genome amplification events in *Y. pestis*, identify their main drivers and analyze the gene content of the amplified regions.

Methods and Algorithms: During this study, 11 *Y. pestis* strains were sequenced using both Illumina and Oxford Nanopore platforms. A new pipeline was developed for the detection of amplified regions based on raw reads and annotation of the potential amplification drivers. In addition to our data, more than 1000 *Y. pestis* read archives were downloaded from the SRA database and processed by the new pipeline. Potential amplification drivers were identified using ICEscan, barrnap and blastn tools.

Results: This study was initiated when we observed a few coverage anomalies in the genome assemblies of 11 sequenced *Y. pestis* strains. Genomes of strains 1627, 1728, 1815, and 1818 had large chromosome regions with local coverage depth significantly higher than the median chromosome coverage depth. The common feature of these regions was flanking them by a pair of identical IS-elements that indicates that the copy number increase could be caused by either RecA-dependent site-specific recombination or "copy-and-paste" transposition [4]. Based on long reads, we inferred that each of these over-coverage events had a structure of tandem genome amplification. Families of IS-elements triggering the recombination in our genomes were annotated as IS3 and IS200/IS605. Since these IS-families are most represented in *Y. pestis* genomes, we could not conclude if there are the preferred amplification drivers, so we decided to expand the search space and downloaded more than one hundred raw read archives from SRA marked as *Y. pestis* whole-genome sequencing data. Firstly, we processed about 100 randomly selected read archives manually and found that over-coverage events appear in at least half of sequenced *Y. pestis* isolates. In rare cases, the amplified regions

were flanked by rRNA gene clusters, which means that these tandem amplifications were more likely the results of RecA-dependent recombination than transposition. Next, we processed all collected data using a new developed pipeline that included the detection of amplified regions and identification of potential amplification drivers.

We found that there are no preferred types of IS-elements or any other repeated sequences triggering the amplification. In most cases, amplified regions were reproducible only between samples belonging to one bioprojet but a few regions were reproducible between different bioprojects. Thus, the 46 kb amplified region that we detected in our 1627 strain was found in *Y. pestis* PBM19. Interestingly, the genome assembly available in RefSeq for this strain does not contain the tandem repeat despite usage of long reads.

After the amplified regions were identified and annotated, we performed the gene function enrichment to check if there are metabolic pathways that are preferably over-covered. We did not find significant over-representation of individual pathways in the analyzed regions so the amplified regions seem to be randomly located on the chromosome and depend only on the presence of repeat sequences. The length and the copy number increase rate of the amplified regions were diverse: the longest identified fragments had the length higher than 1 Mb, and the highest number of tandem repeats was more than 20x. Expectedly, higher copy number was mostly observed for shorter regions that is likely explained by change in fitness cost [5]. Independently, the regions with significantly lowered coverage depth were considered. The only reproducible under-covered regions were *pgm*, and *Y. pestis* filamentous phage [6]. According to the published results, *pgm* is known to be occasionally eliminated under laboratory conditions in curtain *Y. pestis* strains [7], while the filamentous phage is stably integrated into the chromosome only in 1.ORI strains and almost exclusively extrachromosomal in other *Y. pestis* strains [8].

We did not find clear evidence of stable tandem amplification in bacteria different from *Y. pestis* since all described amplification events were associated with stress conditions. Even species close to *Y. pestis* such as *Y. pseudotuberculosis* and *Y. enterocolitica* do not have in their genomes such number of short repeats and the only described case of tandem duplication in them is an amplification of 28 kb fragment containing *blaA* gene in *Y. enterocolitica* appeared under antibiotic stress [9]. Thus, it makes it more intriguing to next investigate the mechanisms and the adaptive role of such copy number dynamics in the plague pathogen.

Funding: This work did not receive any financial support from commercial or non-commercial organizations.

References

1. Andersson D.I., Hughes D. Gene amplification and adaptive evolution in bacteria. *Annu Rev Gen.* 2009;43:167-195
2. Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G. et al. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature*. 2001;413(6855):523-527
3. Bochkareva O.O., Dranenko N.O., Ocheredko E.S., Kanevsky G.M., Lozinsky Y.N. et al. Genome rearrangements and phylogeny reconstruction in Yersinia pestis. *Peer J*. 2018;6:e4545
4. Reams A.B., Roth J.R. Mechanisms of gene duplication and amplification. *Cold Spring Harbor Perspect Biol*. 2015;7(2):a016592
5. Ravatn R., Studer S., Springael D., Zehnder A.J., van der Meer J.R. Chromosomal integration, tandem amplification, and deamplification in Pseudomonas putida F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from Pseudomonas sp. strain B13. *J Bacteriol.* 1998;180(17):4360-4369

6. Derbise A., Chenal-Francisque V., Pouillot F., Fayolle C., Prévost M.C. et al. A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. *Mol Microbiol*. 2007;63(4):1145-1157
7. Leal-Balbino T.C., Leal N.C., Nascimento M.G., Oliveira M.B. et al. The pgm locus and pigmentation phenotype in Yersinia pestis. *Gen Mol Biol*. 2006;29:126-131
8. Chouikha I., Charrier L., Filali S. et al. Insights into the infective properties of YpfΦ, the *Yersinia pestis* filamentous phage. *Virology*. 2010;407(1):43-52
9. Seoane A., Sánchez E. et al. Tandem amplification of a 28-kilobase region from the Yersinia enterocolitica chromosome containing the blaA gene. *Antimicrob Agents Chemother*. 2003;47(2):682-688